FOIL
FORUM OF INSURANCE LAWYERS

Informing Progress - Shaping the Future

FOIL Update 5th March 2026

AI-Driven Deception: How Insurers Are Adapting to Deepfakes

AI has always been acknowledged as a great opportunity for insurers, but its increased adoption has progressed discussions on its associated challenges and threats. Of these, few have unsettled the risk landscape as rapidly as deepfakes, with the rapid development of synthetic media giving rise to new forms of fraud or extortion and exposing businesses to attacks with increased speed and scale.

Deepfakes are causing insurers to undertake a fundamental reassessment of how risk is defined, priced and transferred, particularly within cyber, crime and professional liability markets. The challenges extend beyond technology to touch the foundations of trust on which underwriting traditionally relies, with the impacts reshaping coverage structures, claims dynamics and risk mitigation strategies across the sector.

The Acceleration of Risk

The term deepfake relates to audio, videos and images generated or manipulated using AI that are capable of convincingly impersonating the voices, faces or behaviours of real people. Rather than using malware to directly target vulnerabilities in IT systems, criminals use deepfakes to exploit human judgment by pretending to represent a trusted individual or organisation.

The challenge for traditional cyber defences is accurately detecting deepfakes in real-time before any damage or harm is caused. However, studies suggest an overwhelming

proportion of consumers are unable to identify deepfakes, while state-of-the-art detection systems only report accuracy levels of between 55-70% in real-world applications.

Data published by human risk management specialist Keepnet Labs shows that the volume of deepfake files is growing exponentially, rising from approximately 500,000 in 2023 to more than 8 million by 2025. This represents an increase of more than 1,500% in just two years, with projections suggesting deepfake content will increase by 900% annually.

The challenge is amplified by the lines between what is real and fake being blurred like never before, with technology evolving from niche, novelty AI into operational cybercrime. Fraudsters are utilising cloned voices, fabricated video calls and false documentation to bypass traditional controls. Social engineering attacks, which have historically been reliant on phishing emails, increasingly exploit real-time audio and video impersonation.

The insurance market data reflects this shift, with AI-driven deepfake activity contributing to a 3x year-on-year increase in social engineering incidents between 2023 and 2024 and a significant rise in related cyber insurance claims in 2024.

A Unique Challenge for Insurers

Traditional cyber risks, such as ransomware or data breaches, typically leave forensic traces that enable detection, analysis and investigation. Deepfakes, by contrast, represent a significant shift in digital manipulation by weaponising authenticity to make false audio, video or image evidence appear genuine.

The example of UK engineering group Arup serves as a warning, where losses of approximately £20 million followed the fraudulent cloning of a senior executive in a video conference. The fraud used psychology and sophisticated deepfake technology to gain confidence and trick employees into making genuine transactions; no systems were compromised, and no data was lost. By the time fraud was detected, the funds were lost.

From an underwriting perspective, this creates three problems:

- **Attribution –** the increasing complexity in identifying the source and authenticity of information to distinguish between genuine instruction and deepfake manipulation.

- **Aggregation risk** - the growing availability and use of AI tools enable scalable attacks across multiple insureds at the same time.

- **Coverage -** many legacy policies were never drafted with synthetic media in mind and lack the explicit language to address the exposures, which creates ambiguity.

Legal analysis shows that deepfake incidents now simultaneously impact multiple policy areas, including cyber, commercial crime, D&O, media liability and professional indemnity cover. As a result, insurers face mounting debates over which policy should respond.

The Emerging Coverage Gap

Throughout 2024 and 2025, several cyber insurance carriers began rewriting policy language to exclude losses involving AI-generated deepfake fraud from standard social engineering coverage. These exclusions often target synthetic media used in impersonation or automated fraud schemes, and the practice has become informally known as 'coverage drift', where evolving technology exposes gaps between insured expectations and contractual wording.

Social engineering endorsements have historically been associated with fraudulent emails or written instructions and use relatively straightforward language. Deepfakes, however, extend deception into live audio and video, meaning policy definitions must account for real-time impersonation across collaboration platforms and messaging tools. The risk for policyholders is that, depending on the interpretation of policy wording, an incident may bear the characteristics of cyber fraud but fall outside both cyber and crime coverage.

From Transfer to Active Defence

Rather than relying solely on exclusions, parts of the cyber insurance market are moving towards proactive adaptation. In late 2025, cyber insurer Coalition introduced a global Deepfake Response Endorsement, extending cyber policies with specialised coverage to address reputational damage, fraud response and technical investigation following synthetic media attacks.

The endorsement provides access to specialist forensic analysis, legal support for content removal and assistance with crisis communications, signalling a transition in cyber insurance from financial reimbursement to enhanced incident response capability.

This approach reflects a wider transformation sometimes described as active insurance, where insurers leverage data and insight to act as risk partners rather than merely a passive payer of claims. Such developments mark a significant turning point in the industry and indicate that deepfake exposure is recognised as a board-level risk requiring dedicated insurance solutions.

Underwriting Response

The underwriting reaction to the growing threat of synthetic media increasingly focuses on resilience rather than just prevention. Businesses seeking favourable terms on cyber coverage are now expected to demonstrate capabilities in several areas, including:

- Training employees on recognising synthetic media;
- Embedding secure payment verification procedures;
- Introducing mandatory multi-factor authentication;

- Defining incident response plans to address incidents of AI-driven deception;
- Implementation of deepfake detection technologies.

Organisations adopting similar controls can reduce the likelihood of incidents and perhaps improve their negotiating position during policy placement. At the same time, insurers are investing heavily in analytics and AI-based detection systems to validate claims evidence and identify manipulated media.

There is, however, an ongoing challenge between generative AI capabilities and verification tools, which impacts detection, as fraudsters continuously adapt their techniques to avoid automated checks.

Pricing and Systemic Risk Concerns

Developments in generative AI have made deepfakes inexpensive to produce and globally scalable, enabling untrained individuals to create deepfake media from simple text prompts. This means a single fraudulent campaign is able to target hundreds of insured organisations simultaneously using publicly available data on executives and senior staff. As realism improves, insurers are shifting focus from isolated claims to correlated losses where a single, widespread event can cause multiple policies to fail at the same time.

Recent industry analysis points to insurers now facing systemic exposure as deepfakes grow increasingly realistic and difficult to detect, raising familiar questions previously seen in cyber catastrophe modelling: whether losses can be diversified, if premiums are adequate for AI-scaled fraud and whether government-backed reinsurance mechanisms should be introduced. The latter is increasingly seen as necessary due to the rise of large-scale perils, such as cyber warfare and climate change, that impact all society and extend beyond the capacity of private markets.

Regulatory and Legal Implications

Regulators are also intensifying their scrutiny of deepfakes in recognition of their role as a hybrid threat that blurs the boundaries between fraud, misinformation, privacy breaches and cyberattacks. Their versatility allows deepfakes to be used for several malicious purposes at once, complicating legal classification and compelling governments to adopt comprehensive legal frameworks to address AI-generated, non-consensual content.

A number of key questions have emerged across jurisdictions that need to be tackled to determine what developments are necessary, notably whether AI impersonation constitutes identity theft, how liability is assigned when synthetic content spreads online and what duty of care falls on organisations to verify digital communications. As legal standards develop, insurers will inevitably adjust coverage triggers and exclusions to align with evolving judicial interpretation.

Insuring Trust

Deepfakes have emerged as a significant threat to the insurance industry, fundamentally challenging the established 'seeing is believing' principle that supports claims processing, underwriting and risk assessment, and undermining the assumption that evidence can be trusted. In a society where voices, faces and documents can be fabricated with near perfection, verification is as important as indemnification.

The insurance sector has responded with a combination of adaptability and caution by introducing a combination of clearer policy wording, specialist endorsements and preventative solutions. Indications are that further developments may include dedicated AI liability insurance, mandatory authentication standards linked to coverage, industry-wide data sharing and expanded collaboration between insurers and technology providers. As synthetic media continues to mature, insurers are underwriting authenticity alongside cyber risk.

Deepfakes are more than just another cyber threat vector; they redefine how deception operates in digital economies by targeting human trust rather than technological weaknesses. Cyber insurers already feel the impact through rising claims, but AI-generated manipulation is expanding fraud exposure and testing traditional coverage boundaries across multiple business lines.